The Economics of Reducing Health Risk from Food

EDITED BY
Julie A. Caswell

Proceedings of NE-165 Conference
June 6-7, 1995
Washington, D.C.

PART ONE: Choosing Strategies for Health Risk Reduction

2. Prevention and Treatment in Food Safety: An Analysis of Conceptual Issues

James Barrett and Kathleen Segerson

Food Marketing Policy Center
Department of Agricultural and Resource Economics
University of Connecticut

Prevention and Treatment in Food Safety: An Analysis of Conceptual Issues

James Barrett (jpb94001@uconnvm.uconn.edu)

Kathleen Segerson (segerson@uconnvm.uconn.edu)

Department of Economics
University of Connecticut, Storrs, CT 06269-1063

Keywords: Foodborne illness, damage functions, optimal regulation

2

Prevention and Treatment in Food Safety: An Analysis of Conceptual Issues

James Barrett and Kathleen Segerson[1]

The main goal of food safety policy is to reduce the damages that would result from ingestion of contaminated food. Prevention of contamination is one approach to reducing potential damages. The majority of policies governing food safety have focused on preventing foodborne illnesses. The U.S. Food and Drug Administration (FDA) and U.S. Department of Agriculture (USDA) regulate many food production processes and monitor much of the output from the food industry in an attempt to prevent harmful products from reaching the market. Environmental Protection Agency (EPA) regulations governing the use of pesticides on food are aimed at not allowing more than a certain number of additional cases of illness (usually on the order of one additional case per million per lifetime; see National Academy of Sciences (1987) for a more thorough explanation of EPA regulations). These regulations are extensive and highly specific.

A second approach to reducing damages is through treatment of illnesses when they occur. Unlike prevention, the question of treatment of illnesses has received very little attention from government agencies. Yet the theoretical literature suggests that it may be more efficient to respond to illnesses ex post with medical treatment rather than attempting to prevent them ex ante, because the responses can be individually tailored to suit the nature of the problem and because response is only required if illness occurs.[2] If society is risk neutral and treatment and prevention are equally effective at reducing damages caused by illness, then treatment would always be preferred. However, in general the two approaches will not be equally effective in reducing damages. If prevention is more effective, then the use of some preventative methods may be efficient. The question then becomes: What is the efficient combination of prevention and treatment to use in response to the problem of food safety?

There exists very little literature that addresses the potential tradeoff between prevention and treatment. Using a first-order Markov process, Heffley (1982) develops a decision rule for allocating expenditures between prevention and treatment in the health care context. However, in his model, the choice of prevention and treatment is made simultaneously, so that the level of prevention taken has no effect on the level of treatment needed, and the ex ante vs. ex post nature of the problem is not fully examined. Polinsky and Shavell (1994) develop a model that examines the choice of accident prevention and damage mitigation in the production of some good. (Oil spills and toxic waste leakages are cited as examples.) Their focus is on how different liability rules affect the producer's behavior, and they do not consider the possibility of a limited budget. They focus instead on inducing the firm to employ the optimal levels of care and mitigation, which are taken as given; they do not examine the interaction between the choice of prevention and mitigation. Raucher (1986) presents a theoretical framework for measuring the costs and benefits of groundwater protection. However, the model assumes that the government always chooses "the economically appropriate" response to incidences of contamination. Prevention decisions are made with treatment decisions taken as given, so that there is no need to choose between the two. Although all these papers examine situations in which treatment and prevention decisions must be made, they do not consider the efficiency implications of choosing ex ante vs. ex post variables.

This chapter uses an economic framework to determine the extent to which limited resources should be invested in treatment and in prevention and investigates some of the factors that influence the efficient levels of each. We consider three alternative social criteria by which to judge the efficiency of the investment in prevention and treatment. The first is standard Pareto optimality,[3] or the minimization of total social costs, where those costs include both the costs of prevention and treatment and the damages from food contamination that might result despite the investments in prevention and treatment. The second criterion is expected expenditure minimization subject to a maximum damage constraint. This corresponds to a situation in which the authorities have determined some minimum level of safety (or maximum level of expected damages) and the objective is to achieve this level at a minimum cost to society.[4] The third criterion is expected damage minimization subject to a resource constraint. This is relevant in a situation in which a planner has a limited budget available and, given that budget, the objective is to minimize the social damages due to foodborne illnesses.

The results of our model suggest that, while the allocation of resources between prevention and treatment under the second two criteria will usually differ from the Pareto optimal allocation, they share some of the same properties of the Pareto optimal allocation. Specifically, we show that the allocations under all three criteria lie on a type of expansion path along which the rate that prevention and treatment can be substituted for each other while maintaining a given budget is the same rate at which they can be traded while maintaining a given level of expected damages, which is in turn equal to the rate at which they can be traded while maintaining a given level of expected aggregate costs. Thus, this condition for being on the expansion path provides a rule for allocating resources efficiently that can be used 1) when the regulatory agency faces a budget constraint, 2) when it seeks to ensure that expected damages do not exceed some threshold, and 3) when it is unconstrained. However, the impact of changes in certain factors that influence the efficient allocation depends on which criterion is used. For example, contrary to what might be expected, when the probability of contamination occurring rises exogenously, the efficient levels of both prevention and treatment may fall under a resource constraint, but both may rise under a damage constraint.

In other cases, our results suggest that the optimal levels of prevention and treatment move in opposite directions in response to an exogenous parameter change. This suggests a potential tradeoff between the two approaches to reducing damages. For some, the notion of reducing expenditures on illness prevention in order to increase expenditures on illness treatment presents a distasteful choice; reducing prevention expenditures will lead to a higher incidence of illness and thus more human suffering. However, every dollar that goes to prevention cannot go towards treatment of illnesses that have occurred. This reduction in treatment is also an increase in human suffering, so that the choice being made is not between dollars and human suffering, but rather between two alternative means of reducing suffering. By allowing some additional illnesses to occur, it may be possible to treat more illnesses once they do occur; if this is true, then it may be more efficient to allocate more resources towards treatment.

Nonetheless, we recognize that there may be some minimum level of safety which society wants to maintain in its food supply; providing less than that level of safety (by allowing expected damages to go above a certain level) may be a disservice to society. The inclusion of the second criterion, social wealth maximization subject to a maximum damage constraint, is an attempt to address this concern. In addition, the consideration of interim damages in the comparative-static section is an attempt to address the fact that some portion of damages may not be treatable.

An Overview of the Model

Consider the following situation: There is some food that is being produced and consumed which may or may not cause the consumer to become ill. The producer can undertake some degree of prevention that reduces the probability of the consumer becoming ill. If the consumer does become ill, he can seek out treatment to mitigate the damages being suffered. For simplicity, it is assumed that the consumer can take no preventative action to reduce the probability of illness. Additionally, the effectiveness of prevention and of treatment are assumed to be known with certainty. It is also assumed that both the producer and consumer are risk neutral or that actuarially fair insurance is available and that all damages are monetarily compensable. Thus, expected utility can be maximized by maximizing expected net wealth.

Pareto Optimality

This section will explore the familiar efficiency criterion of maximizing one individual's expected utility subject to the constraint that the other individual's expected utility not fall below some specified level. Given risk neutrality, maximizing expected utility is equivalent to maximizing expected net wealth. Thus, the problem is to

$$\text{(1)} \qquad \underset{T,A,s}{\textit{Maximize}}\ W_I^O - A - s$$

$$s.t.\ W_V^O - P(A)T - P(A)D(T) + s \geq \bar{W}_V,$$

where:

W_I^O	is the producer's initial wealth (I for injurer),
W_V^O	is the consumer's initial wealth (V for victim),
$\bar{W}_V$	is some predetermined target level of expected wealth,
A	is the amount of prevention taken by the producer, with units chosen so that the price of A is one,
P	is the probability that the consumer gets ill,
T	is the amount of treatment for damages sought out by the consumer, with units chosen so that the price of T is one,
D	is the damages suffered by the consumer, with
$D' < 0$	so that each additional unit of T reduces damages, but
$D'' > 0$	so that it reduces damages less as it is used more,
$P' < 0$	so that each additional unit of A reduces the probability of illness, and
$P'' > 0$	so that it reduces the probability less as it is used more, and
s	is an exogenous lump-sum transfer from the producer to the consumer.

We assume that the consumer has already decided how much of the product to consume. Thus, we do not model the consumption decision. In a more general model where consumption is endogenous, the probability that the consumer gets ill could depend on his consumption decision, inter alia.[5] The lump-sum transfer, s, is required to separate the distributional effects of the Pareto criteria from its efficiency goal. It can be set to ensure that the constraint holds with equality,[6] i.e.:

(2) $$s = \bar{W}_V - W_V^O + PT + PD.$$

Given s and the exogeneity of the initial and target wealth levels, the problem in (1) is equivalent to minimizing total expected social costs. Thus, (1) becomes:

(3) $$\underset{A,T}{Minimize}\ A + P(A)T + P(A)D(T).$$

The first-order conditions for T and A are:

(4) $$-D' = 1$$

and

(5) $$-P' \cdot (T + D) = 1,$$

respectively. These simply require that the marginal benefit of each is equal to its marginal cost. The marginal benefit of prevention is the reduced probability that treatment will be needed and damages will be suffered. The marginal benefit of treatment is simply the reduction in damages associated with an increase in treatment. Units are chosen so that the marginal cost of each is 1. Note that under this criterion, the optimal level of T is determined solely by (4) and is independent of the probability of an illness. Thus, the treatment decision is separable from the prevention decision.

In Figure 2.1, the point EC^* represents the minimum level of total expected costs EC = A + P(A)T + P(A)D(T) that could be incurred by society. Higher levels of total expected costs are represented by concentric circles or ovals (drawn as circles for simplicity) spreading out from EC^*. The further from EC^* an iso-expected-cost-circle lies, the higher the level of expected costs it represents.

The solution to the Pareto criterion (A^*, T^*) yields EC^* and the associated level of expected damages $P(A^*)D(T^*)$ represented by the iso-expected-damage curve D^*. Also drawn is the iso-expenditure curve for $R^* = A^* + P(A^*)T^*$. It should be noted that at EC^* the slope of the iso-expected-damage curve is equal to the slope of the iso-expenditure curve. This can be demonstrated by combining equations (4) and (5) to obtain the marginal condition:

(6) $$\frac{1 + P'T}{P} = \frac{P'D}{PD'}.$$

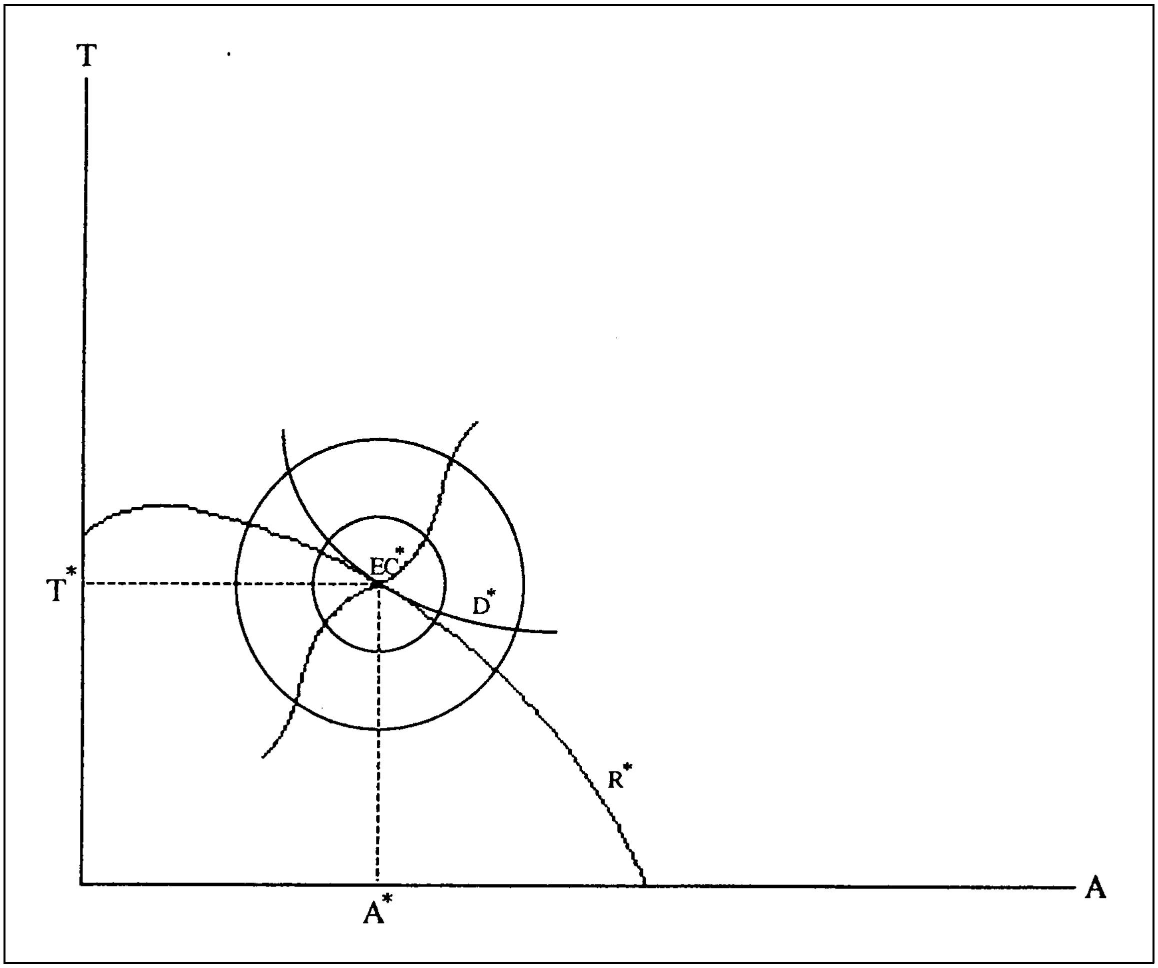
Thus, at the Pareto optimal levels, expenditure is minimized given the associated level of expected damages. Likewise, expected damages are minimized given the level of expected expenditure. It is possible to draw an expansion path through all the points where the slope of an iso-expected-damage curve,

$$-\frac{P'D}{PD'},$$

is equal to the slope of the iso-expenditure curve,

$$-\frac{1 + P'T}{P}.$$

FIGURE 2.1 Solution of the Pareto Optimal Problem



As shown below, the efficient levels of prevention and treatment under the second two criteria lie on this expansion path as well.

Maximum Damage Constraint

The remaining two criteria replace the Pareto optimality criterion with a second-best criterion. In this case, the objective is to minimize total expected expenditure subject to a constraint on the level of damages expected to be suffered by the consumer. The objective function is:

(7a)
$$\underset{T,A}{Minimize}\ A + P(A)T$$
$$s.t.\ \ P(A)D(T) \le \bar{D}.$$

It can be easily shown that (7a) is equivalent to[7]

(7b)
$$\underset{A,T}{Minimize}\ A + P(A)T + P(A)D(T)$$

$$s.t.\ P(A)D(T) \leq \bar{D}.$$

The equivalence of the two problems implies that minimizing expected expenditure subject to an expected damage constraint is, in fact, a constrained version of the first-best Pareto optimal problem in (3). It therefore yields second-best levels of prevention and treatment that will in general yield a higher level of expected total social costs than the first-best Pareto optimal levels. Thus, while choosing to maintain a minimum level of health, or maximum level of expected damages, will reduce one component of total expected social costs, namely, expected damages P(A)D(T), it will necessarily increase the other component, expected expenditure A + P(A)T. The fact that (7) is a constrained version of (3) implies that the increase in expected expenditure will always be at least as great as and generally greater than the decrease in expected damages, thereby increasing expected total social costs.

If the constraint in (7) holds with equality,[8] the first order conditions for the optimal values of A and T under the damage-constrained problem are:

(8)
$$\bar{D} - PD = 0,$$

(9)
$$1 + P'T - \lambda P'D = 0, \text{ and}$$

(10)
$$P - \lambda PD' = 0,$$

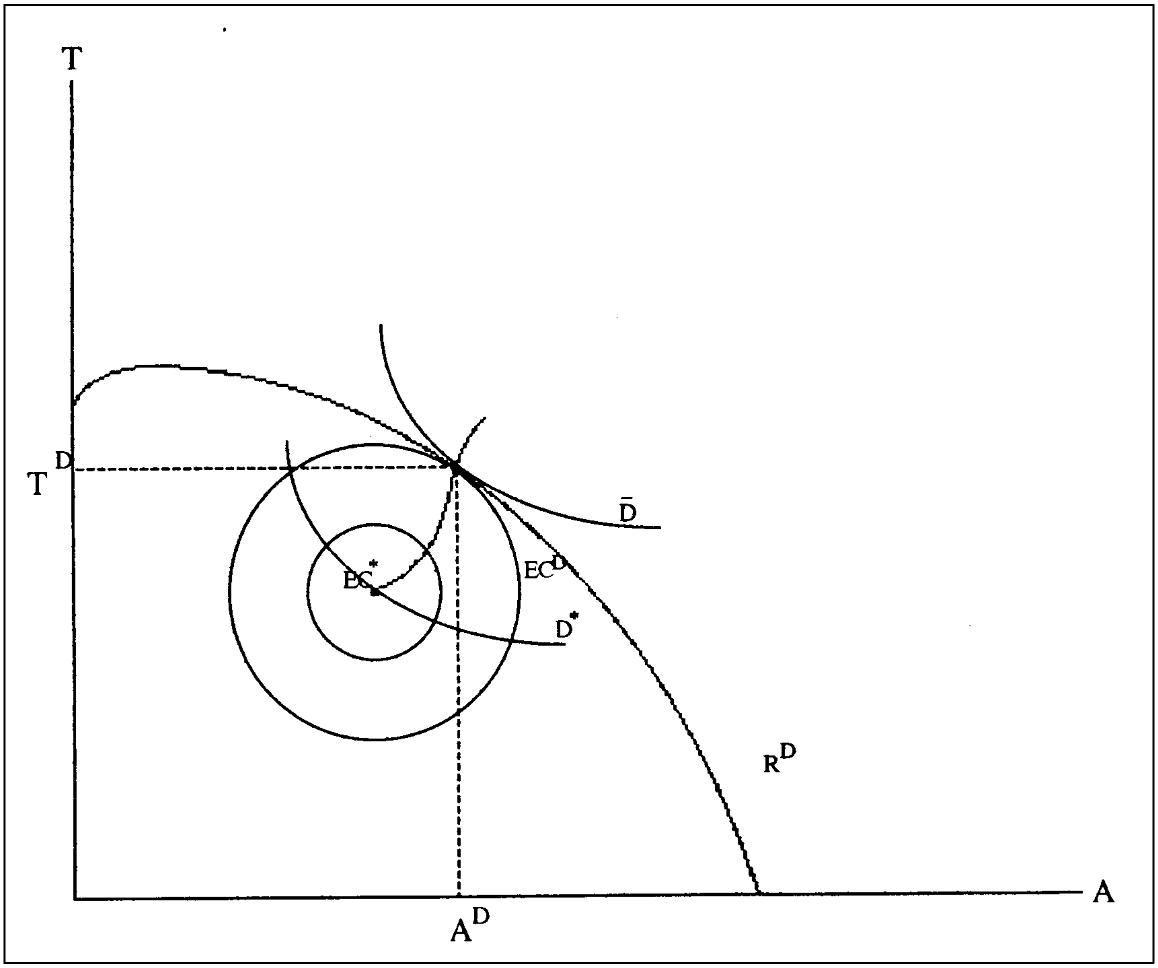
where λ is the Lagrangian multiplier on the constraint. The conditions (9) and (10) are clearly similar to the conditions (4) and (5) from the Pareto criterion and would be identical if λ were equal to -1. Combining (9) and (10) yields the marginal condition:

(11)
$$\frac{1 + P'T}{P} = \frac{P'D}{PD'},$$

which simply states that in equilibrium the slope of the iso-expenditure curve (left-hand side) is equal to the slope of the iso-damage curve (right-hand side). This marginal condition was also implied by the first-order conditions for the Pareto criterion. Thus, as noted above, Pareto optimality implies that expected expenditure is minimized subject to the associated level of expected damages. However, minimizing expected expenditure subject to a given level of expected damages does not imply Pareto optimality.[9] In this sense, Pareto optimality is a stronger criterion than is the efficiency criterion used in the damage-constrained problem. Note also that under the damage-constrained criterion, the decision regarding the level of treatment is no longer separable from the prevention decision. Specifically, unlike under the Pareto optimal criterion, the efficient level of treatment depends on the probability of an illness occurring. This is because the constraint requires that an increase in prevention or treatment requires an offsetting decrease in the other, forcing them to be traded off for one another at a rate which keeps expected damages constant.

Figure 2.2 illustrates graphically the solution to the damage-constrained problem. Here, the level of expected damages is constrained to be $\bar{D}$ which we assume is lower than D^*.[10] Graphically, the efficient levels of A and T are the levels that are on the lowest (i.e., inner most) iso-expected cost curve given the constraint. The solution is (A^D,T^D), which requires that R^D resources be allocated to the problem and results in expected total costs of EC^D. An expansion path can also be drawn to show the

FIGURE 2.2 Solution of the Damage-Constrained Problem



different equilibrium points associated with different levels of the constraint. This expansion path is simply the upper half of the expansion path from the Pareto problem.

Resource Constraint

In this case, the objective is to minimize expected damages subject to a constraint on the amount of resources that can be spent on prevention and treatment. Note that at the time the expenditure on prevention is made, it is not known whether or not an illness will occur and thus whether or not treatment will be required. We therefore specify the resource constraint in terms of the actual expenditure on prevention plus the expected expenditure on treatment.[11] The problem is then to

(12a)
$$\underset{A,T}{Minimize}\ P(A)D(T)$$
$$s.t.\ A + P(A)T \leq \bar{R}.$$

Again, it can be shown that this problem is equivalent to:

(12b) $$\underset{A,T}{Minimize}\ P(A)D(T) + A + P(A)T$$

$$s.t.\ A + P(A)T \leq \bar{R}.$$

Thus, as with the damage-constrained problem, minimizing expected damages subject to a resource or budget constraint is equivalent to a constrained version of the Pareto optimality problem. It thus yields second-best levels of prevention and treatment that generate higher expected total social costs than occur under Pareto optimality. While the expected expenditure on prevention and treatment will be smaller, the decrease in this component of social costs will be offset by the resulting increase in the expected damages. Thus, as with the damage-constrained problem, expected total social costs will increase.

If the constraint holds with equality, the first-order conditions for the resource-constrained problem are:

(13) $$\bar{R} - A + PT = 0,$$

(14) $$P'D - \gamma \cdot (1 + P'T) = 0, \text{ and}$$

(15) $$PD' - \gamma P = 0,$$

where γ is the Lagrangian multiplier on the constraint. Again, this would be identical to the Pareto solution if the multiplier were equal to -1. Combining (14) and (15) again yields the marginal condition:

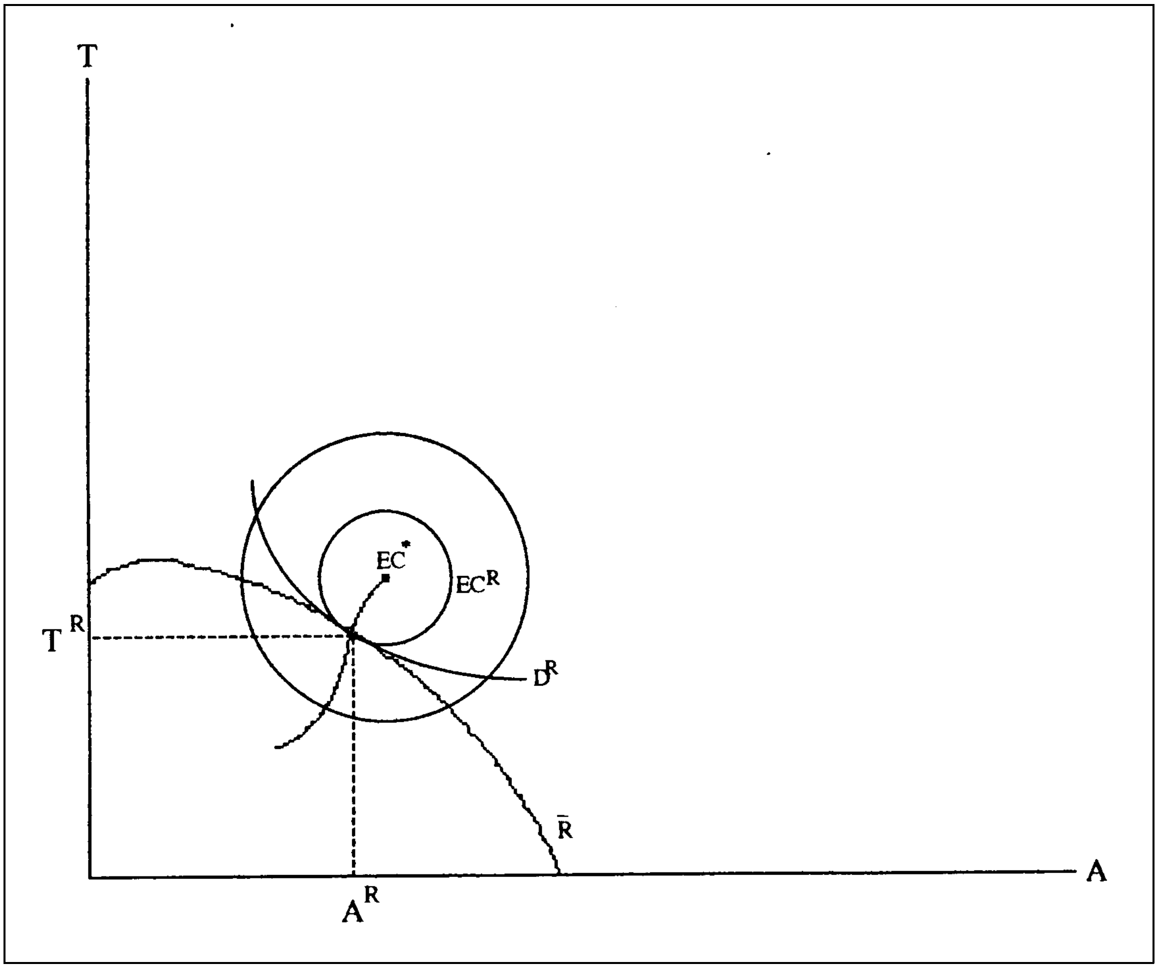
(16) $$\frac{1 + P'T}{P} = \frac{P'D}{PD'},$$

which is equivalent to (11). This reflects the fact that (7a) and (12a) are duals of each other. Thus, in order to minimize expected damages subject to an expected expenditure constraint, it is necessary to minimize expected expenditure subject to the corresponding expected damage constraint. As with the damage-constrained problem, the optimal level of treatment is not independent of the level of prevention. Specifically, it depends on the probability of illness since the constraint forces prevention and treatment to be traded off for one another to keep the level of expected expenditures constant.

Figure 2.3 illustrates graphically the solution to the resource-constrained problem. Here, the amount of resources available is constrained to be less than or equal to $\bar{R}$ (which must be below R^* if the constraint is binding). The efficient allocation of resources again is the allocation that lies on the lowest possible (i.e., inner most) iso-expected cost curve given the constraint. The solution is (A^R, T^R), which results in expected damages of D^R and leaves society with total expected costs of EC^R. An expansion path is again drawn. In this case it is the lower half of the expansion path from the Pareto problem.

The relationship between the three problems should now be fairly clear. The second two criteria are simply constrained versions of the first, unconstrained problem, and produce a family of solutions of which the Pareto solution is a special case. This family of solutions is the set of all combinations of A and T which satisfy (11) (or, equivalently, (16)). Additionally, each one of these tangency points corresponds to a certain level of expected costs, and it can be shown that when this condition is satisfied, the slope of the iso-expected-cost,

FIGURE 2.3 Solution of the Resource-Constrained Problem



$$-\frac{1 + P'T + P'D}{P + PD'},$$

will also be equal to the slopes of the iso-expenditure curve and the iso-expected damage curve. The set of points where these three curves are tangent forms the expansion path drawn in Figure 2.1. The unconstrained Pareto problem yields a unique solution, the single point on this expansion path that minimizes total expected costs. The other two problems yield families of solutions that are on this expansion path. The exact point on the expansion path is determined by the constraint. In the special cases where $\overline{D}$ equals D^* for the damage-constrained problem and where $\overline{R}$ equals R^* for the resource-constrained problem, the two constrained problems both yield the unconstrained Pareto solution.

Comparative Static Results

Having characterized the efficient levels of prevention and treatment under three alternative efficiency criteria, we turn next to a consideration of some of the factors that affect those choices. The first is the level of interim damages. Since time often passes between the onset of symptoms and the beginning of treatment, some damages are suffered in this interim. The severity of these interim damages can be expected to alter the efficient allocation between prevention and treatment. For example, one might expect that the greater the interim damages would be, the more we would want to invest in prevention rather than treatment. The model confirms this intuition in that it calls for an increase in prevention, but the effects on treatment are less clear and depend on which criterion is used. Second, the likelihood that an illness would occur given some level of prevention is expected to affect the choice of prevention and treatment as well. Specifically, one might expect that the more likely the illness is, the more we would want to invest in preventing it. However, the model does not necessarily support this intuition, and the results again are sensitive to the criterion examined. This section will examine how these two factors affect the efficient levels of prevention and treatment. It is particularly interesting to note that while these two second best problems are duals of each other, they can yield different comparative static results, some of which may seem counterintuitive at first glance.

Interim Damages

When people become ill, there is often a lag between the time that the symptoms first manifest themselves and the time that successful treatment begins. In the meantime, some damages are suffered. The severity of these interim damages can be influenced by any number of things, such as economic or psychological reluctance of the sick person to see a doctor promptly, or the virulence of an illness that causes severe damages even over a small period of time. One would normally expect that the more severe interim damages are, the more prevention and the less treatment would be employed. However, whether or not this intuition is borne out depends on which efficiency criterion is used.

A simple modification to capture the effects of interim damages under Pareto optimality yields the objective function:

(17) $$\underset{A,T}{\text{Minimize}}\; A + P(A)T + P(A)(\mathbf{D} + D(T))$$

where $\mathbf{D}$ is interim damages. This formulation assumes that the interim damages simply shift the damage function without altering the effectiveness of treatment in reducing damages. The first-order conditions for prevention and treatment are:

(18) $$1 + P' \cdot (T + \mathbf{D} + D(T)) = 0 \text{ and}$$

(19) $$D' + 1 = 0,$$

respectively.

Using these first-order conditions, we derive the comparative-static results

$$\frac{\partial A^*}{\partial \mathbf{D}} = \frac{-P'PD'}{|H_1|} \geq 0 \quad\text{and}\quad \frac{\partial T^*}{\partial \mathbf{D}} = \frac{0}{|H_1|} = 0,$$

where $|H_1| \geq 0$. These results follow clearly from the first-order conditions. An increase in interim damages increases the damages suffered if illness occurs, so that the marginal benefit of reducing the probability of an accident increases, and the equilibrium level of prevention therefore rises. Interim damages have no effect on the marginal benefit of treatment, so that when they increase, there is no effect on the equilibrium treatment level.

Modifying the damage-constrained model results in the objective function:

(20)
$$\underset{A,T}{Minimize}\ A + P(A)T$$

$$s.t.\ P(A)(\dot{D} + D(T)) \leq \bar{D}.$$

If the constraint is binding, this yields first-order conditions:

(21)
$$\bar{D} - P(A)[\dot{D} + D(T)] = 0,$$

(22)
$$(1 + P'T) - \lambda P' \cdot (\dot{D} + D) = 0, \text{ and}$$

(23)
$$P - \lambda PD = 0,$$

where again λ is the Lagrangian multiplier. Given that the objective function is convex in A and T, it is a relatively simple matter to obtain comparative static results,

$$\frac{\partial A^D}{\partial \dot{D}} = \frac{-P^2 P\lambda D''(\dot{D} + D) - \lambda P'(PD')^2}{|H_2|} \geq 0 \text{ and}$$

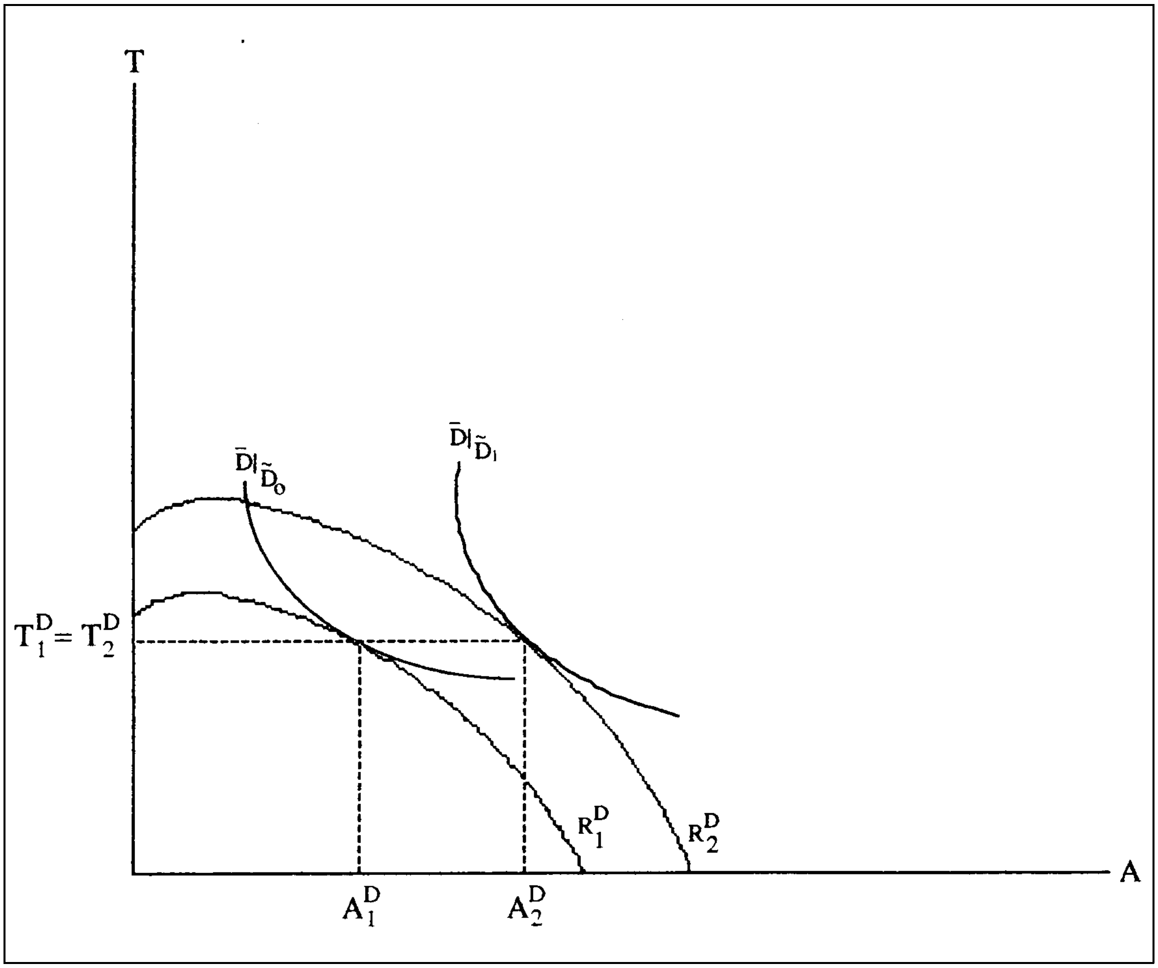
$$\frac{\partial T^D}{\partial \dot{D}} = \frac{D'P[P'^2\lambda \cdot (\dot{D} + D) + PP''(T - \lambda \cdot (\dot{D} + D))]}{|H_2|}$$

which is indeterminate, where $|H_2| \leq 0$. Thus, under this efficiency criterion an increase in interim damages has the expected result of increasing investment in prevention. However, the effect on treatment is indeterminate.

Figure 2.4 shows the effects of an increase in interim damages on the damage-constrained problem. The inner iso-expenditure curve and iso-expected-damage curve represent the original case, and the outer pair represent what happens when interim damages increase. Two different effects can be identified. The first is the damage-shift effect. The increase in interim damages shifts the iso-expected damage curve outward, requiring an increase in expenditures to keep the constraint satisfied. (Both iso-expected-damage curves represent the same level of expected damages.) This moves us to a new iso-expenditure curve, R_2^D, which is parallel to the original iso-expenditure curve, R_1^D. The second effect is the damage-substitution effect. The increase in interim damages causes the iso-expected-damage curve to become steeper. This occurs because reducing the probability of illness through increased prevention now reduces the probability of suffering a greater level of damages than before. It would take a greater amount of treatment to match the reduction in expected damages, thereby increasing the marginal rate of substitution and making the iso-expected-damage curve steeper.

Both of these effects tend to increase prevention, and it is shown to rise unambiguously from A_1^D to A_2^D. However, the effects work in opposite directions on treatment; the damage-shift effect tends to increase treatment, and the damage-substitution effect tends to decrease it. It is impossible to say which effect would dominate without making additional, somewhat arbitrary, assumptions. The diagram is drawn such that treatment does not change, T_1^D equals T_2^D, but it should be clear that either an increase or decrease is possible.

FIGURE 2.4 Effect of an Increase in Interim Damages on the Damage-Constrained Problem



In the resource-constrained case, the problem is now to:

(24)
$$\underset{A,T}{\textit{Minimize}}\ P(A)[\tilde{D} + D(T)]$$
$$s.t.\ A + P(A)T \le \bar{R}.$$

This yields the first-order conditions:

(25)
$$\bar{R} - A - P(A)T = 0,$$

(26)
$$P'[\tilde{D} + D(T)] - \gamma \cdot (1 + P'T) = 0, \text{ and}$$

(27)
$$PD' - \gamma P = 0,$$

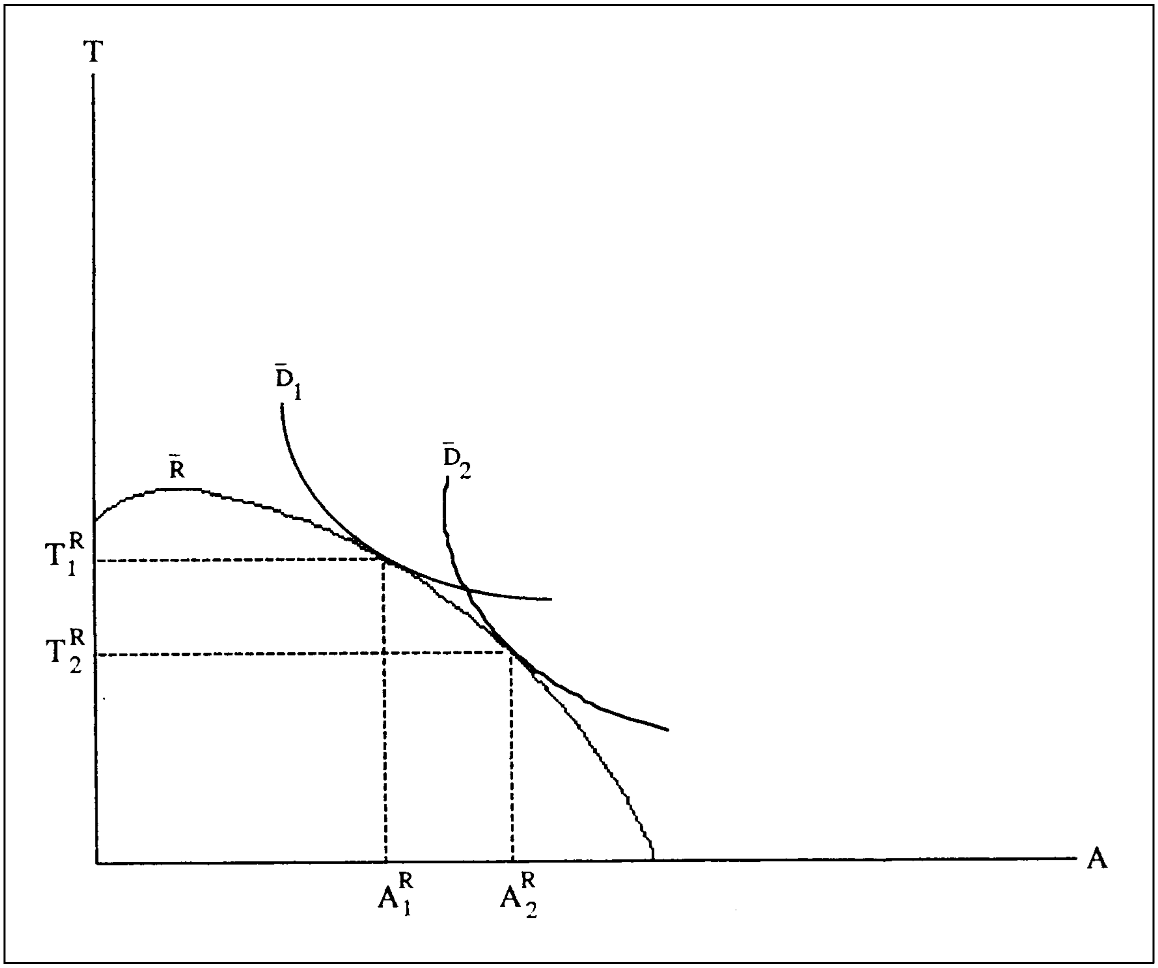
where γ again is the Lagrangian multiplier. This time, however, (assuming convexity) the comparative static results are consistent with a priori intuition:

$$\frac{\partial A^R}{\partial \bar{D}} = \frac{P'P^2}{|H_3|} \geq 0 \text{ and } \frac{\partial T^R}{\partial \bar{D}} = \frac{P'P(1 + P'T)}{|H_3|} \leq 0,$$

where $|H_3| \leq 0$.

Figure 2.5 shows the effect of an increase in interim damages in the resource-constrained problem. In this case, the increase in interim damages does not shift the constraint. However, as before, there is a damage-substitution effect, which causes the iso-expected-damage curves to become steeper. However, the fact that expenditures cannot go up in this case means that the level of expected damages must rise. The second iso-expected-damage curve, D_2^R, represents a higher level of expected damages than the original, D_1^R. (Recall that higher iso-expected-damage curves represent lower levels of expected damages.) The damage-substitution effect makes D_2^R steeper than D_1^R, and, as in the last case, tends to decrease treatment and increase prevention. Without a shift in the constraint to mitigate the damage effect on T, it is clearly shown that prevention increases and treatment decreases unambiguously in this case.

FIGURE 2.5 Effect of an Increase in Interim Damages on the Problem



Change in Probability

One would expect the probability of any given individual becoming ill to depend on factors other than the amount of prevention taken to avoid getting ill. In the context of food safety, exogenous parameters can be expected to play a role in the probability of illness. For example, weather could affect the probability of a *Salmonella* or other bacteriological illness occurring, as warm weather is more conducive to bacterial growth. This section examines how an exogenous increase in the probability of illness affects the choice of prevention and treatment. One might expect that such a change would tend to decrease treatment and increase prevention because increasing the probability of illness increases the probability that the treatment will actually have to be employed which is akin to raising the price of treatment. Again, however, whether or not this occurs depends on which criterion is used.

Modifying the Pareto optimal problem gives the unconstrained objective function:

$$\underset{A,T}{Minimize}\ A + P(\alpha,A)T + P(\alpha,A)D(T), \tag{28}$$

where α is the exogenous parameter that affects the probability of illness. We define α such that the partial derivative, P_α, is positive. This yields first-order conditions:

$$1 + P_A(T + D) = 0 \text{ and} \tag{29}$$

$$1 + D' = 0. \tag{30}$$

Using these conditions, we derive the comparative-static results:

$$\frac{\partial A^*}{\partial \alpha} = \frac{-P_{A\alpha}(T + D)D''}{|H_4|},$$

the sign of which is the opposite of the sign of the cross-partial $P_{\alpha A}$, and

$$\frac{\partial T^*}{\partial \alpha} = \frac{0}{|H_4|} = 0,$$

where $|H_4| \geq 0$. Again, the results are fairly intuitive. Because α only affects the probability of an accident and treatment is chosen without regard to the probability, a change in α will not affect the choice of treatment. The effect of a change in α on the choice of prevention depends on the sign of $P_{\alpha A}$. If $P_{\alpha A}$ is negative, then the increase in α has made prevention more effective in reducing the probability of an accident, and more prevention should be used to take advantage of this. If $P_{\alpha A}$ is positive, then prevention has become less effective, and less should be used. If $P_{\alpha A}$ is equal to zero, then the increase in α has had no effect on the effectiveness of prevention, and the equilibrium level of prevention does not change. Since it seems more likely that an increase in α would reduce the effectiveness of prevention, the rest of the comparative-static results will be based on the assumption that $P_{\alpha A}$ is positive.

Modifying the damage-constrained version of the model gives the problem:

$$\underset{A,T}{Minimize}\ A + P(\alpha,A)T \tag{31}$$

$$s.t.\ P(\alpha,A)D(T) \leq \bar{D}.$$

The solution to this problem gives the first-order conditions:

(32) $$\bar{D} - P(\alpha,A)D(T) = 0,$$

(33) $$1 + P_A T - \lambda P_A D = 0, \text{ and}$$

(34) $$P - \lambda P D' = 0,$$

where λ again is the Lagrangian multiplier. Assuming convexity of the objective function, we obtain the following comparative-static results:

$$\frac{\partial A^D}{\partial \alpha} = \frac{-P_\alpha D^2 P_A P \lambda D'' + P_{\alpha A}(T - \lambda D)(PD')^2}{|H_5|} \text{ and}$$

$$\frac{\partial T^*}{\partial \alpha} = \frac{D(T - \lambda D)(-PD' P_A P_{\alpha A} - P_\alpha P_{AA})}{|H_5|}$$

where $|H_5| \leq 0$. The sign of both of these depend on the cross-partial derivative, $P_{\alpha A}$. If $P_{\alpha A}$ is positive, then

$$\frac{\partial T^D}{\partial \alpha} \geq 0.$$

The sign of

$$\frac{\partial A^D}{\partial \alpha}$$

depends not only on the sign of $P_{\alpha A}$, but also on its size. If $P_{\alpha A}$ is positive and large enough,

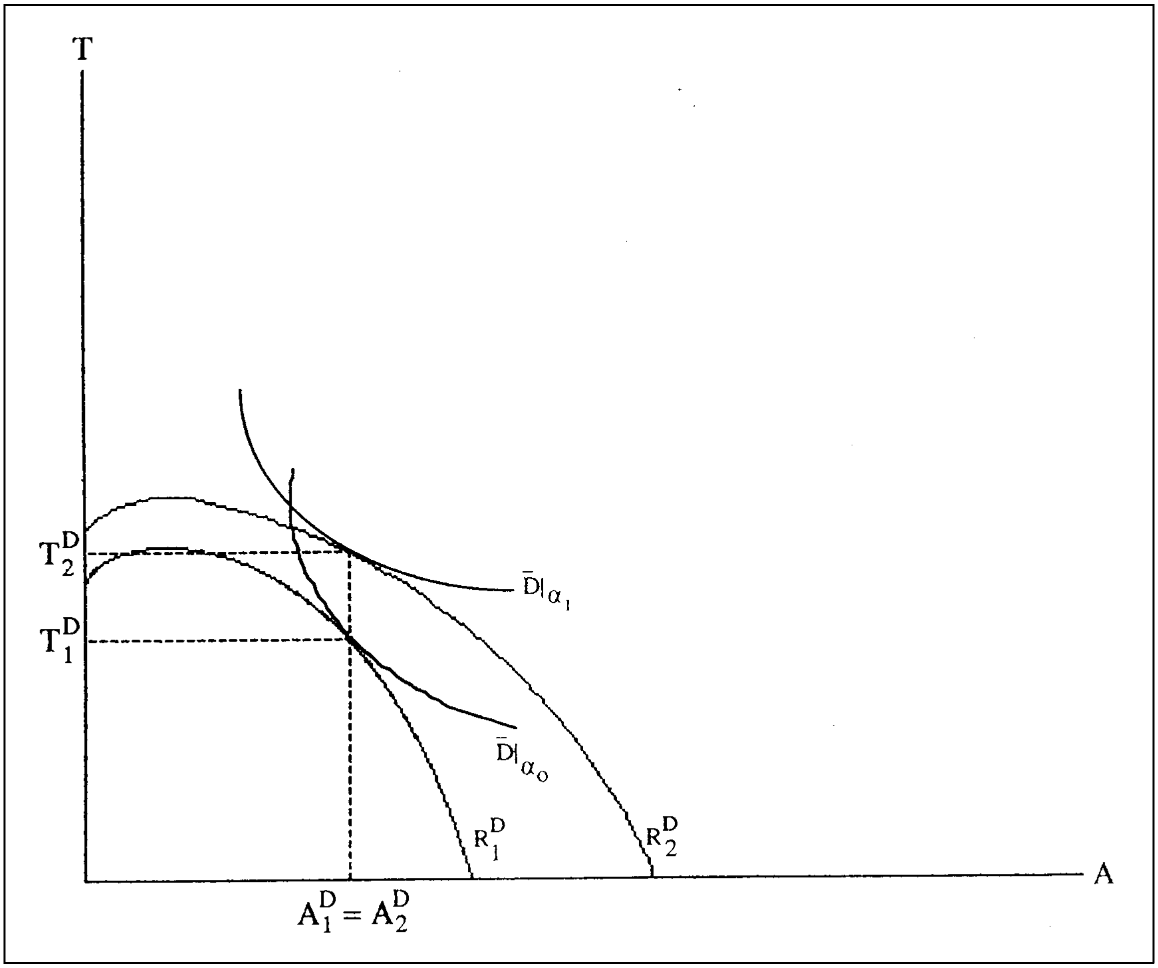
$$\frac{\partial A^D}{\partial \alpha} \leq 0,$$

otherwise

$$\frac{\partial A^D}{\partial \alpha} \geq 0.$$

Figure 2.6 shows the effects of a change in α on the damage-constrained problem, assuming $P_{\alpha A}$ is positive. The inner iso-expenditure curve R_1^D and iso-expected-damage $\bar{D}|_{\alpha_0}$ represent the original case, and the outer pair, R_2^D and $\bar{D}|_{\alpha_1}$, represent the situation after α increases. This time, three different effects can be identified. The first is the damage-shift effect. As was the case with interim damages, the increase in α shifts the iso-expected-damage curve outward so that more resources are required to maintain the same level of expected damages. (Again, the two curves $\bar{D}|_{\alpha_0}$ and $\bar{D}|_{\alpha_1}$ represent the same level of expected damages.) This requires that we move to a higher iso-expenditure curve, R_2^D. This effect tends to increase both prevention and treatment. At the same time, the damage-substitution effect tilts the iso-expected damage curve so that it becomes flatter. This also tends to increase treatment but tends to decrease prevention. The third effect is a resource-substitution effect. The change in α rotates the iso-expenditure curve so that the rate of substitution between prevention and treatment changes. In the absence of a constraint on expected damages, the increase in α would cause the iso-expenditure

FIGURE 2.6 Effect of an Exogenous Increase in the Probability of Illness on the Damage-Constrained Problem



curve to tilt inward, resulting in a higher level of expected damages (a lower iso-expected damage curve). However, the expected damage constraint does not allow that, so that the iso-expenditure curve merely rotates, rather than tilting inward, remaining on the same iso-expected damage curve. The direction of the resource-substitution effect depends on which way and how much the iso-expenditure curve rotates. This is determined by the size of $P_{\alpha A}$. If $P_{\alpha A}$ is large, the iso-expenditure curve will get steeper, and the resource-substitution effect will add to the damage-substitution effect's tendency to decrease prevention and increase treatment. If $P_{\alpha A}$ is small (or negative), the iso-expenditure curve gets flatter as it rotates, tending to decrease treatment and increase prevention. If this flattening effect were large enough, then the combination of the resource-substitution effect and the damage-shift effect would dominate the damage-substitution effect's tendency to decrease prevention, and prevention would increase. Otherwise, the damage-substitution effect would dominate, and A^D will decrease.

In the diagram, the second iso-expenditure curve is shown to be flatter than the original, but only flat enough so that prevention doesn't change ($P_{\alpha A}$ positive and small). It should be easy to see that if it were to get even flatter, prevention would go up, and if it were to get any less flat or steeper than the original iso-expenditure curve, then prevention would go down.

Under the resource-constrained criterion, the problem becomes:

(35)
$$\underset{T,A}{Minimize}\ P(\alpha,A)D(T)$$
$$s.t.\ A + P(\alpha,A)T \le \bar{R}.$$

This yields the first-order conditions:

(36)
$$\bar{R} - A - P(\alpha,A)T = 0,$$

(37)
$$P_A D - \gamma \cdot (1 + P_A T) = 0, \text{ and}$$

(38)
$$PD' - \gamma P = 0,$$

and the comparative static:

$$\frac{\partial A^R}{\partial \alpha} = \frac{P_\alpha TPD''(1 + P_A T) + P^2 P_{\alpha A}(D - \lambda T)}{|H_6|} \text{ and}$$

$$\frac{\partial T^R}{\partial \alpha} = \frac{-(D - \lambda T)(PP_{\alpha A}(1 + P_A T) + P_\alpha TP_{AA})}{|H_6|}$$

where $|H_6| \le 0$. As in the damage-constrained problem, the signs of the comparative statics depend on the size and sign of the cross-partial $P_{\alpha A}$. If it is positive and large,

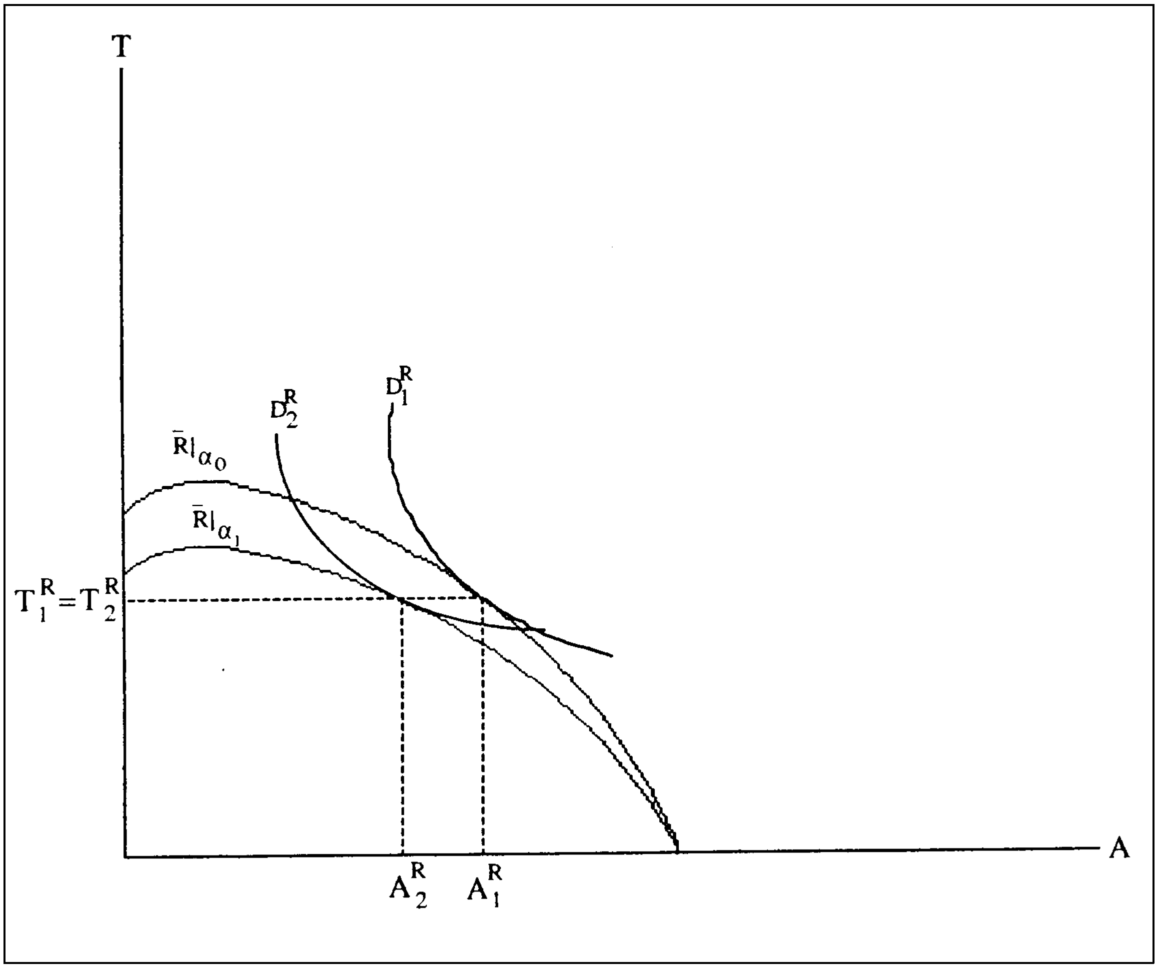
$$\frac{\partial A^R}{\partial \alpha} \le 0 \quad \text{and} \quad \frac{\partial T^R}{\partial \alpha} \le 0.$$

If it is positive and small,

$$\frac{\partial A^R}{\partial \alpha} \le 0 \quad \text{still but} \quad \frac{\partial T^R}{\partial \alpha} \ge 0.$$

Figure 2.7 shows the effect of an increase in α on prevention and treatment in the resource-constrained problem. The outer iso-expenditure curve $\bar{R}|_{\alpha_0}$ and iso-expected-damage curve D_1^R represent the situation before the change in α, and the inner pair $\bar{R}|_{\alpha_1}$ and D_2^R, represent the situation after the change, where $\bar{R}|_{\alpha_0}$ and $\bar{R}|_{\alpha_1}$ represent the same level of expenditures. Again, three effects can be identified. First, the resource-substitution effect again rotates the iso-expenditure curve, this time flattening it unambiguously. Unlike the previous case, because the level of expected-damages is not constrained, the resource-substitution effect is accompanied by a resource-shift effect. This is analogous to the income effect in standard microeconomic theory, as the increase in P raises the expected price of treatment. The resulting decrease in purchasing power makes both prevention and treatment less affordable, so that they both tend to decrease. Since treatment is now more likely to be needed, the same actual levels of prevention and treatment result in a higher expected level of expenditures, so that a reduction in both is required to satisfy the resource constraint. The combination of these two effects results in the new iso-expenditure curve. At the same time, the damage-substitution effect flattens

FIGURE 2.7 Effect of an Exogenous Increase in the Probability of Illness on the Resource-Constrained Problem



out the iso-expected-damage curves, so that D_2^R is flatter than D_1^R. This adds to the resource-shift effect's tendency to reduce prevention, but moves treatment in the opposite direction, tending to increase it. If the resource-substitution effect is large enough, when combined with the resource-shift effect, it will dominate the damage-substitution effect's tendency to increase treatment, and treatment will fall. If not, the damage-substitution effect will dominate, and T^R will increase. The diagram is drawn so that treatment does not change, but it should be clear that a larger rotation of the iso-expenditure curve will cause T^R to fall, and a smaller rotation will cause T^R to rise. There is no damage-shift effect in this scenario because the resource constraint keeps expenditures from increasing, so that there is no outward shift in the iso-expenditure curve.

Uncertainty

Many elements of the food safety problem are characterized by uncertainty. One important source of uncertainty is the effectiveness of the means used to reduce the expected damages caused by foodborne illness. Using a method first introduced by Lichtenberg and Zilberman (1988) and further developed by Lichtenberg et al. (1989), this section examines how uncertainty regarding the effectiveness of treatment in reducing damages affects the efficient allocation of resources between prevention and treatment. This "safety rule" approach allows for "uncertainty-compensated tradeoffs between risk and social cost ... reflecting both uncertainty about risk and decision-makers' preferences regarding that uncertainty" (Lichtenberg et al. 1989). Under this approach, the policymaker sets a risk standard and a safety margin, ρ, expressed as a percentage. The object is to ensure that the risk standard is violated no more than $1 - \rho$ percent of the time at a minimum cost to society.

In the original model by Lichtenberg and Zilberman (1988), and in subsequent applications by Lichtenberg et al. (1989) and Harper and Zilberman (1992), the random variable of interest is the probability that a randomly selected individual will experience some adverse health effect of a fixed magnitude; there is no possibility for treatment or remediation. Additionally, their models consider several parameters of which overall risk is a multiplicative combination. The assumption is that these parameters are distributed lognormally so that the log of health risk is the sum of the logs of the parameters. In the problem considered here, the random variable is the level of expected damages. The distribution of this random variable depends on the level of treatment. Instead of a probabilistic risk standard, the policymaker chooses an expected damages standard, $\bar{D}$; actual expected damages should exceed $\bar{D}$ not more than $1 - \rho$ percent of the time. Rather than consider several sources of uncertainty, here we focus only on uncertainty regarding the effectiveness of treatment in reducing damages. As a result, there is no need to assume a lognormal distribution, and we assume instead that the damage function is distributed normally about some mean. Additionally, the original models and applications assume that some parameters affect mean risk more than the variance of risk and that the opposite is true of other parameters. Here, we assume that the variance of damages, σ^2, is constant, and that only the mean of damages, $\mu(T)$, is affected by treatment.[12]

Formally, in the context of our model, the safety rule approach can be expressed as:

$$Pr\{P(A) \cdot D(T) \geq \bar{D}\} \leq 1 - \rho. \tag{39}$$

Because damages are normally distributed, it is possible to express this in terms of the mean and standard deviation of damages:

$$P(A)[\mu(T) + Z(\rho) \cdot \sigma] \leq \bar{D}, \tag{40}$$

where $Z(\rho)$ is the critical value of the normal distribution that is exceeded with probability $1 - \rho$.

Rather than minimizing social costs subject to a constraint determined by the safety rule, as was done in the original safety rule models, the model here will focus only on the resource constrained problem. Assuming that (39) holds with equality, we consider two different objective functions that are minimized with and without the resource constraint. The first objective is to minimize the probability that the given standard, $\bar{D}$, is violated, and the other is to minimize the standard (making it as stringent as possible) which will be violated $1 - \bar{\rho}$ percent of the time, where $\bar{\rho}$ is taken as given. We then examine how an exogenous change in the standard deviation of damages, σ, affects the allocation of resources between prevention and treatment. Again, some results confirm our initial intuition while others do not, and the results are shown to depend upon which criterion is used.

Minimizing the Standard

As mentioned above, in this formulation, the objective is to find the most stringent standard that can be expected to be satisfied a given percentage of the time. This may seem the opposite of how standards are set in real life situations, but this formulation is effective in demonstrating the sensitivity of the comparative static results to model specification. The unconstrained problem is to

(41) $$\underset{A,T}{Minimize}\ D = P(A)[\mu(T) + Z(\bar{\rho}) \cdot \sigma].$$

This yields the first-order conditions:

(42) $$P' \cdot (\mu + Z\sigma) = 0 \Rightarrow P' = 0 \text{ and}$$

(43) $$P\mu' = 0 \Rightarrow \mu' = 0.$$

These simply require that prevention and treatment be used until the probability and mean damages of an illness can be reduced no further. It should be clear that this must be the case. In the absence of any cost considerations in the objective function or through a constraint, the marginal cost of both prevention and treatment is zero, and they will both be employed until their marginal benefits are also zero. It should also be clear that the variance of damages has no effect on the equilibrium levels of prevention and treatment.

With a resource constraint, the problem becomes:

(44) $$\underset{A,T}{Minimize}\ D = P(A)[\mu(T) + Z(\bar{\rho}) \cdot \sigma]$$

$$s.t.\ A + P(A) \cdot T \leq \bar{R}.$$

Assuming that the constraint holds with equality (and suppressing arguments for ease of exposition), this yields first-order conditions:

(45) $$\bar{R} - A - P \cdot T = 0,$$

(46) $$P' \cdot (\mu + Z \cdot \sigma) - \lambda(1 + P' \cdot T) = 0, \text{ and}$$

(47) $$P \cdot \mu' - \lambda \cdot P = 0,$$

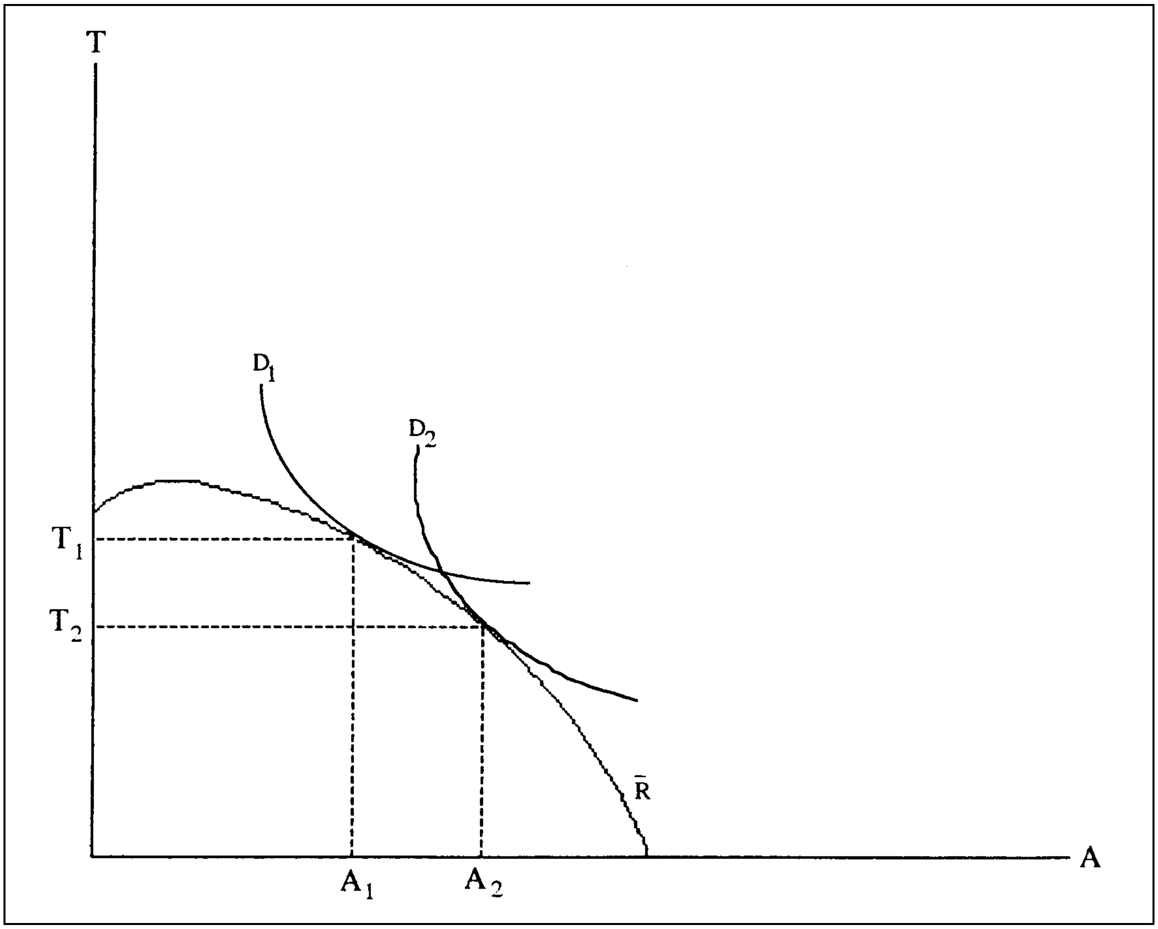
where λ is the Lagrangian multiplier. These yield the comparative static results

$$\frac{\partial A}{\partial \sigma} = \frac{P^2 ZP'}{|H_7|} \geq 0 \quad \text{and} \quad \frac{\partial T}{\partial \sigma} = \frac{-PZP'(1 + P'T)}{|H_7|} \leq 0,$$

where $|H_7| \leq 0$.

Figure 2.8 shows how an increase in the standard deviation affects the choice between prevention and treatment. The iso-expenditure curve is drawn and labeled $\bar{R}$. As in the resource-constrained version of the interim damage case, only one effect exists. Because the change in σ does not affect the relative "prices" of prevention and treatment, there is no resource-substitution or resource-shift effect.[13] Because the constraint on resources means that the iso-expenditure curve does not shift outward, there is no damage-shift effect either. The only effect that still exists is the damage-substitution effect which makes the iso-expected-damage curve steeper, rotating from D_1 to D_2, so that equilibrium treatment falls from T_1 to T_2 and equilibrium prevention rises from A_1 to A_2. This is consistent with intuition in that we would expect an increase in the uncertainty regarding the usefulness of one of our tools to reduce the extent to which it is employed. As the effectiveness of treatment becomes more uncertain, we would expect that society would be more reluctant to rely on it, thus shifting resources toward prevention and away from treatment.

FIGURE 2.8 Effect of an Increase in the Variance of Damages Given a Resource Constraint and a Fixed Margin of Safety



Minimizing the Probability of Exceeding the Standard

In this scenario, the policymakers have chosen some minimum level of expected damages to serve as the safety standard, and the objective is to choose the levels of prevention and treatment that minimize the probability that the standard is violated. The objective function is to:

(48)
$$\underset{A,T}{\text{Minimize}}\ \rho = Z^{-1}\left(\frac{\bar{D} - P(A)\cdot\mu(T)}{P(A)\cdot\sigma}\right)$$

$$s.t.\ A + P(A)\cdot T \leq \bar{R},$$

or, equivalently,

(49)
$$\underset{A,T}{\text{Minimize}}\ \frac{\bar{D}}{P(A)} - \mu(T)$$

$$s.t.\ A + P(A)\cdot T \leq \bar{R}.$$

It is clear from (49) that under this criterion,

$$\frac{\partial A}{\partial \sigma} = 0 \quad \text{and} \quad \frac{\partial T}{\partial \sigma} = 0.$$

Thus, with this formulation, an increase in uncertainty regarding the effectiveness of treatment has no effect on the amount of resources invested in prevention or treatment.

These results are less intuitive at first glance. Figure 2.9 shows the distribution of expected damages in the absence of any prevention or treatment and the distribution after prevention and treatment are employed (but before σ increases). Prevention and treatment are used to shift the distribution of expected damages downward (to the left) by decreasing the probability of an illness occurring and by reducing the mean of damages if an illness does occur. This has the effect of reducing the area of the distribution which lies above the expected damage standard, $\overline{D}$. Figure 2.10 shows the left most distribution from Figure 2.9 before and after σ increases. The increase in σ widens the distribution, so that a greater area of the distribution is above $\overline{D}$. However, prevention and treatment can do no further good in reducing this area, because they have already shifted the distribution as far left as possible, given the resource constraint.

FIGURE 2.9 Distribution of Expected Damages Before and After Treatment and Prevention Are Employed

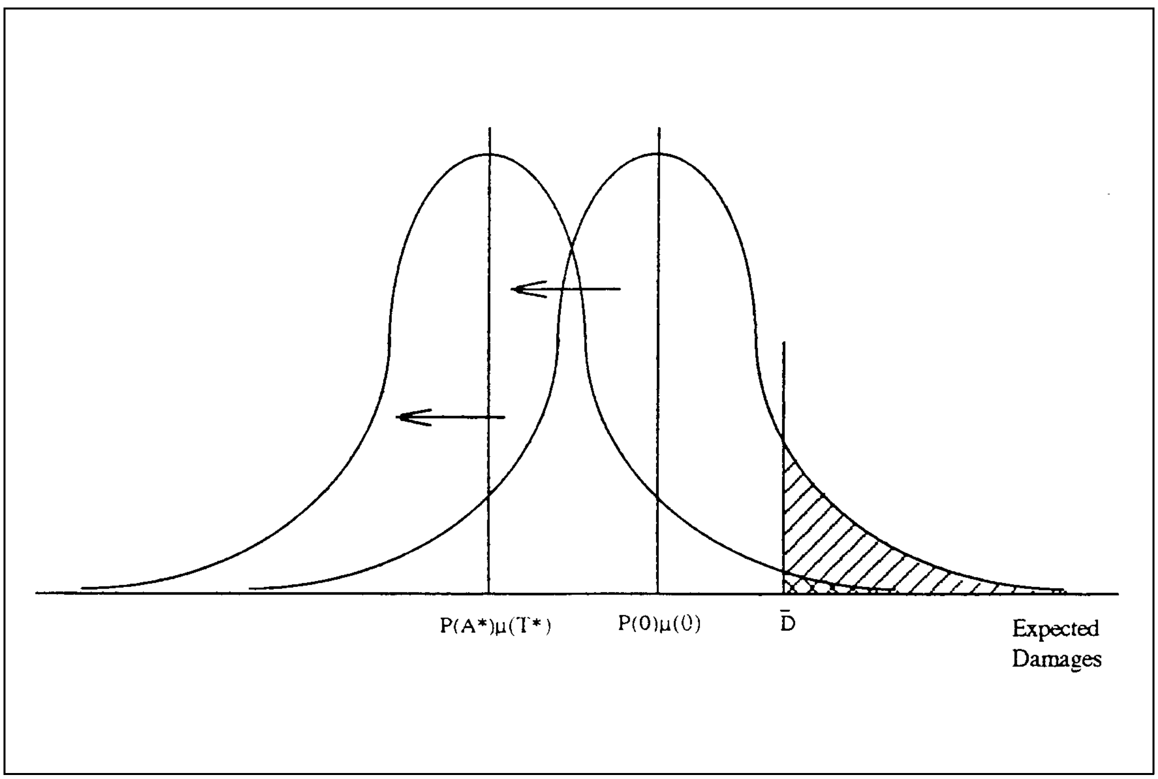
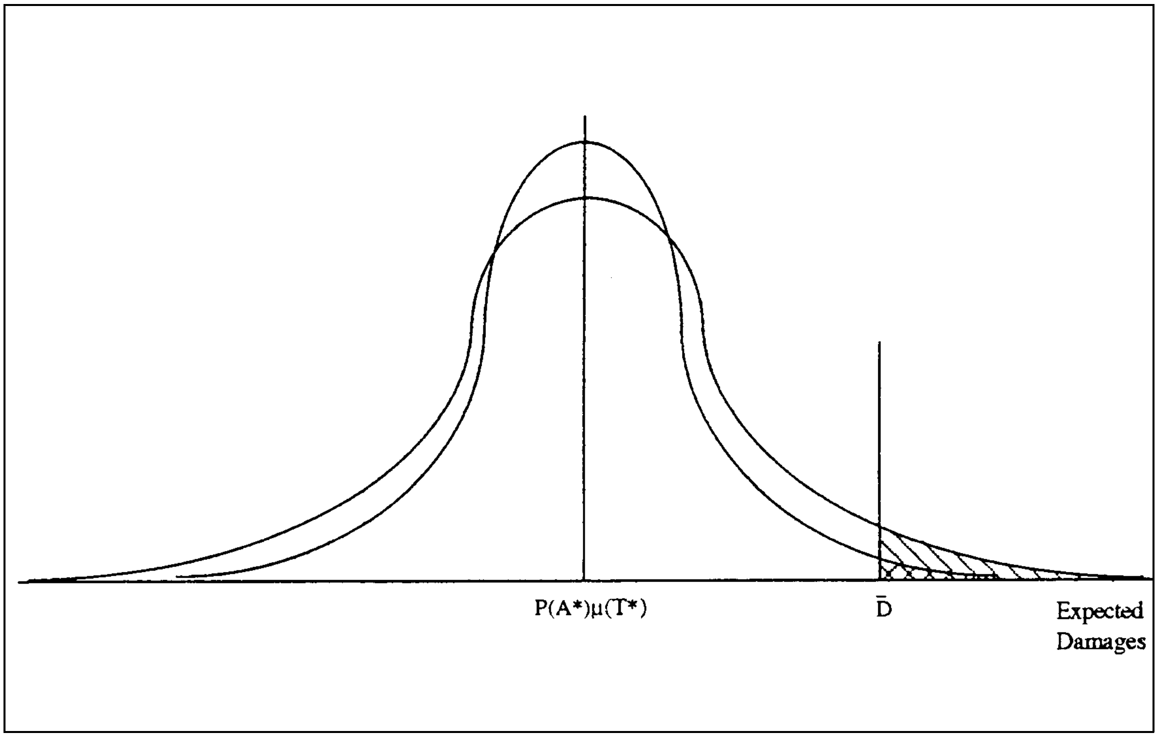


FIGURE 2.10 Effect of an Increase in the Variance of Damages Given a Resource Constraint and a Fixed Level of Expected Food Safety



Conclusions

While theory suggests that the use of treatment may be a useful means of dealing with the problem of food safety, it has received very little attention from regulators and theoreticians alike. If society were risk averse and prevention and treatment were equally effective at reducing expected damages, treatment would be preferred. In general, we do not expect that they will be equally effective, and some combination of the two is preferred. This leads us to ask: What is the optimal combination of prevention and treatment, and what factors affect this combination?

In an attempt to answer these questions, we began by trying first to determine what is the appropriate definition of efficiency in the food safety context. We find that the three seemingly different criteria considered here are in fact closely related, and each yields a set of solutions that are merely subsets of a larger family of solutions. All three criteria yield solutions on a common locus of solutions; the exact location of the solution on the locus depends on which criterion is being examined. The resource-constrained and damage-constrained definitions are constrained versions of the Pareto definition and thus provide second-best solutions, of which the Pareto definition is a special case. Turning to the factors which affect the optimal allocation of resources between prevention and treatment, we find, despite the similarities of the definitions, that the effects of these exogenous factors depend on which definition is used. While the results may seem to contradict intuition at first, they do confirm microeconomic theory.

Our results suggest, on the one hand, that more attention should be paid to the tradeoff between prevention and treatment and that more attention should be focused on the use of treatment as a policy tool for reducing the damages caused by foodborne illness. On the other hand, our results show that the definition of efficiency against which prevention and treatment allocations are judged may play a more important role than is generally thought. Efficiency is often defined without much thought regarding alternative definitions or the fact that these alternatives might yield different, sometimes opposite, conclusions. In situations where ex ante and ex post variables are substituted for one another, the definition of efficiency plays a pivotal role in how allocations are viewed.

Notes

[1]James Barrett is a Ph.D. Candidate, Department of Economics, University of Connecticut and Kathleen Segerson is Associate Professor, Department of Economics, University of Connecticut. The authors wish to thank the Food Marketing Policy Center at the University of Connecticut for funding support of this project.

[2]For a seminal work covering this and other issues, see Raiffa 1968.

[3]Given linear utility functions (constant marginal utilities) and the availability of exogenous wealth transfer mechanisms, Pareto optimality is equivalent to aggregate wealth maximization. See Miceli and Segerson (1995) and Shavell (1987) for a thorough discussion of these issues.

[4]For other examples of this type of formulation, see Lichtenberg and Zilberman (1988), Lichtenberg et al. (1989), and Harper and Zilberman (1992).

[5]It is also possible to model these choices per unit of consumption or production, which would also abstract from the consumption decision.

[6]See Miceli and Segerson (1995) and Shavell (1987) for a thorough discussion of this and related issues.

[7](7a) and (7b) can be shown to be equivalent by comparing the Lagrangians. The Lagrangian for (7a) is:

$$A + P(A) \cdot T - \lambda[P(A) \cdot D(T)] + \lambda \cdot \bar{D},$$

and the Lagrangian for (7b) is:

$$A + P(A) \cdot T + P(A) \cdot D(T) - \mu[P(A) \cdot D(T)] + \mu \cdot \bar{D},$$

or

$$A + P(A) \cdot T + (1 - \mu)[P(A) \cdot D(T)] + \mu \cdot \bar{D}.$$

A comparison of the two Lagrangians shows that the only difference between them is the value of the multipliers which, of course, will not affect the optimal levels of A and T.

[8]If the constraint is not binding, (7b) reduces to (3).

[9]This is analogous to the result that profit maximization implies cost minimization, but cost minimization does not imply profit maximization.

[10]If $\bar{D} \geq D^*$, the constraint in (7a) is not binding and the solution is identical to the Pareto optimal solution.

[11]An alternative is to specify a constraint that requires the sum of actual expenditure on both (A + T) to be less than or equal to the resources available. However, under such a constraint, resources would be "wasted," i.e., the constraint would not be binding, if no illness occurred and thus no expenditure on treatment was needed.

[12]We make this assumption in order to consider exogenous increases in the level of uncertainty. Alternatively, we could specify $\sigma^2 = \sigma^2(\beta,T)$ where β is an exogenous parameter that shifts σ^2 given T.

[13]We would not expect this to be the case if σ affected P, thus changing the slope of the expenditure line.

References

Harper, Carolyn and David Zilberman. 1992. Pesticides and Worker Safety. *American Journal of Agricultural Economics* 74:68-78.

Heffley, Dennis. 1982. Allocating Health Expenditures to Treatment and Prevention. *Journal of Health Economics* 1:265-290.

Lichtenberg, Erik and David Zilberman. 1988. Efficient Regulation of Environmental Health Risks. *Quarterly Journal of Economics* 103:167-178.

Lichtenberg, Erik, David Zilberman, and Kenneth Bogen. 1989. Regulating Environmental Health Risks Under Uncertainty: Groundwater Contamination in California. *Journal of Environmental Economics and Management* 17:22-24.

Miceli, Thomas and Kathleen Segerson. 1995. Defining Efficient Care: The Role of Income Distribution. *Journal of Legal Studies* 24:189-208.

National Academy of Sciences, National Research Council, Board on Agriculture. 1987. *Regulating Pesticides in Food: The Delany Paradox*. Washington, D.C.: National Academy Press.

Polinsky, A. Mitchell and Steven Shavell. 1994. A Note on Optimal Cleanup and Liability After Environmentally Harmful Discharges. *Research in Law and Economics* 16:17-24.

Raiffa, Howard. 1968. *Decision Analysis*. Reading, MA: Addison-Wesley.

Raucher, Robert. 1986. The Benefits and Costs of Policies Related to Groundwater Contamination. *Land Economics* 62:33-45.

Shavell, Steven. 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.